Environmental Science & Technology

Article

pubs.acs.org/est

Comprehensive Emerging Chemical Discovery: Novel Polyfluorinated Compounds in Lake Michigan Trout

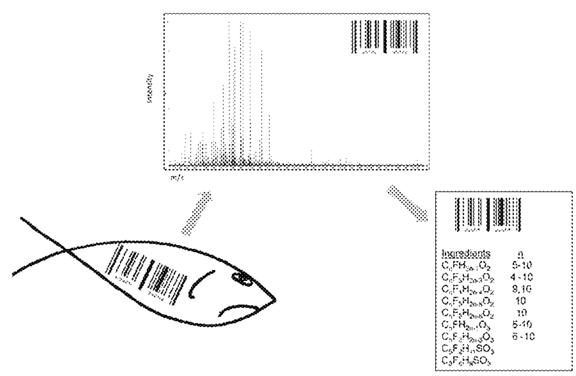
Sadjad Fakouri Baygi,[†] Bernard S. Crimmins,*[,‡,§] Philip K. Hopke,[†,∥] and Thomas M. Holsen[‡]

[†]Department of Chemical and Biochemical Engineering, [‡]Department of Civil and Environmental Engineering, and [∥]Institute for a Sustainable Environment, Clarkson University, 8 Clarkson Avenue, Potsdam, New York 13699, United States

[§]AEACS, LLC, Alliance, Ohio 44601, United States

Supporting Information

ABSTRACT: A versatile screening algorithm capable of efficiently searching liquid chromatographic/mass spectrometric data for unknown compounds has been developed using a combination of open source and generic computing software packages. The script was used to search for select novel polyfluorinated contaminants in Great Lakes fish. However, the framework is applicable whenever full-scan, high-resolution mass spectral and chromatographic data are collected. Target compound classes are defined and a matrix of candidates is generated that includes monoisotopic mass spectral profiles and likely fragmentation pathways. The initial calibration was performed using a standard solution of known linear perfluoroalkyl acids. Once validated, Lake Michigan trout data files were analyzed for polyfluoroalkyl acids using the algorithm referencing 3570 possible compounds including C_4-C_{10} perfluoro- and polyfluoroalkyl, polyfluorochloroalkyl acids and sulfonates, and potential ether forms. The results suggest the presence of 30 polyfluorinated chemical formulas which have not been previously reported in the literature. The identified candidates included mono- to hexafluoroalkyl carboxylic acids, mono- and trifluoroalkyl carboxylic acid ethers, and novel polyfluoroalkyl sulfonates. Candidate species identified in lake trout were qualified using theoretical isotopic profile matching, characteristic fragmentation patterns based on known linear perfluoroalkyl acid (PFAA) fragmentation, and retention time reproducibility among replicate extractions and injections. In addition, the relative retention times of multiple species within a compound class were compared based on theoretical octanol–water partition coefficients.



INTRODUCTION

Perfluorinated and polyfluorinated compounds are ubiquitous and considered persistent, and potentially harmful to the environment.[1,2] Perfluorooctanesulfonate (PFOS) and perfluorooctanoic acid (PFOA) have been identified as persistent organic pollutants in the Stockholm Annex B.[3] Recently, C_8 chemistries have been phased out in North America[4] and a proposal to ban the use of PFOA-like substances in the European Union[5] has been issued. However, significant quantities are still produced in Asia[6] along with alternative compounds such as tetrafluoroheptafluoropropoxy propionic acid (HFPO–DA).[7] Historically, long-chain, linear perfluoroalkyl carboxylic acids (PFAAs) and sulfonates (PFAS) have been monitored in the environment and PFOS and PFOA are typically the dominant contributors to the isomeric distribution.[8,1] Studies since early 2000 have shown the presence of PFAAs and perfluoroalkyl sulfonates (PFSAs) in household dust at 10–100 ng/g,[2,9] and human serum worldwide in the ng/mL range.[2,10]

The physical and chemical properties of the $-CF_2-$ backbone provides significant thermal stability and both water and oil repellency.[2] These characteristics are highly desirable for aviation hydraulic fluids, firefighting foams, pesticides, metal plating, electronic devices, mining, carpets, textiles and upholstery, paper and packaging, coating and coating additives,[11] driving the need to replace regulated substances with novel chemicals with similar properties. Recent raw material degradation and end-user product studies have identified a complex array of perfluorinated compounds including polyfluorinated ethers, *n*-ethyl perfluorooctane sulfonamido acetic acid, and fluorotelomer acids.[12,13]

Typical analytical methods use ultrahigh-performance liquid chromatography (UPLC) equipped with triple quadrupole mass spectrometry (TQ).[2,14] The high sensitivity and the use of multiple precursor/product pairs provide a robust method for targeted analysis. Recently, a quadrupole time-of-flight mass spectrometer (QToF) coupled to a UPLC was used to quantify trace-level (ng L^{-1}) concentrations of polyfluorinated acids.[15] In this method, the authors developed a data independent MS/

Received: March 17, 2016
Revised: July 31, 2016
Accepted: August 5, 2016

MS workflow that continuously collected both precursor and product ions (alternating low and high energy channels) with mass accuracies greater than 0.01 Da. Alternative high-resolution mass spectrometers have been employed for perfluorinated compound analysis including Fourier transform ion cyclotron resonance mass spectrometry (FTICR)[13] and linear ion traps.[16] Both techniques allow for exact mass identification and the use of mass defect filtering analyses to better visualize the data. Liu et al.[16] analyzed wastewater from a fluorochemical manufacturing park in China with an HPLC-LTQ-Orbitrap-MS and detected 36 compounds consisting of polyfluorinated sulfates ($C_nF_{2n+3}H_{n-2}SO_4^-$), chlorine substituted perfluorocarboxylates ($ClC_nF_{2n}CO_2^-$), and hydro substituted perfluorocarboxylates ($HC_nF_{2n}CO_2^-$), but it is unclear whether the observed species were ether cleavages or byproducts from other unidentified polyfluorinated compounds.[16]

High-resolution mass spectral data sets can utilize measured mass defects (difference between the exact and nominal mass) of a species for classification purposes.[13,17−19] This technique has been used to elucidate polyfluoroalkyl compound classes generated from the thermolysis of polychlorotrifluoroethylene (PCTFE),[13] in groundwater impacted by aqueous film-forming fire-fighting foams,[19] and industrial wastewater.[16]

Hilton[20] developed an algorithm utilizing 2-Da isotopic clusters of compounds containing halogens for GC × GC-ToF-MS data. Strynar et al.[21] applied isotopic profile matching in addition to accurate mass determination and successfully identified three classes of mono- and polyether perfluorinated compounds in surface water collected in North Carolina, USA using Time of Flight (TOF) data. Identifications were based on the presence of protonated and sodiated dimer species. The nontypical ionization observed illustrates the need for comprehensive screening strategies in environmental samples. Schymanski et al.[22] performed a comprehensive nontarget analysis of water from River Danube in 18 institutes from 12 European countries using different software packages. In total, they identified 649 unique compounds as targeted, tentative, or nontargeted compounds.

Lake trout have been used as bioindicators of the health of the Great Lakes region for decades as part of the U.S. Environmental Protection Agency's Great Lakes Fish Monitoring and Surveillance Program (GLFMSP).[23−25] Recently, the program has adopted a proactive role in identifying new chemicals of concern by using scanning, high-resolution mass spectrometers to generate searchable chemical fingerprints of biological samples.[26] The current filtration approach utilizes high-resolution mass spec data files converted into a MATLAB compatible format. The search algorithm references a user-generated candidate compound matrix (or library) consisting of a comprehensive linear combination of elements. Boundary conditions for each element type and number are used to detect chemical compound classes (perfluoro-, polyfluoro-, mixed halogenated, acids, ethers, sulfonates) and class configurations (i.e., rings, units of unsaturation, per/polyfluoroalkyl chains). In this work, Lake Michigan lake trout mass spectral data files were explored for novel polyfluorinated compounds, matching experimental spectra with potential candidate compounds. Once identified, a series of tests, including retention time reproducibility and fragmentation, were used to qualify candidate peaks.

■ EXPERIMENTAL SECTION

Sample Preparation. The algorithm was applied to archived lake trout data files generated using a recently developed hybrid targeted/nontargeted method for measuring long-chain perfluorinated acids.[15] A brief description of the sample processing methodology and instrumentation, a Waters Acquity UPLC equipped with a Waters Xevo G2 QToF in MS^e mode, can be found in the Supporting Information.

Candidate Compounds. The search algorithm was initialized using a candidate molecular formula matrix containing species recently discovered from fluoropolymer thermal decomposition,[13] present in industrial wastewater,[16,27] and discharged into the environment.[21] Species considered had the following molecular formula:

$$C_cO_oF_fCl_{cl}H_hS_s$$

The subscripts *c*, *o*, *f*, *cl*, *h*, and *s* indicate the number of each element in the candidate compounds with *c* ranging from 4 to 10, *o* being 2 for carboxylic forms, 3 for carboxylic ether, and sulfonate forms, and 4 for the ether sulfonate form. The summation of variables *f*, *cl*, and *h* was set so that all carbon atoms were fully saturated and the compound was deprotonated (negative ESI mode was used). Therefore, only parent ions ($[M - H]^-$) of the candidate compounds were initially sought. This molecular configuration resulted in 3570 compounds assuming there were no π bonds between carbon atoms as these were deemed to have a low likelihood of being present in these biological samples. The current search method yields a narrow range of the compound classes likely to be present in the sample (Table S.1). However, the candidate lists within a compound class are comprehensive. Additional elements such as N, P, and Br could also be added to the seed molecular formula, but were not included in the current screening.

Candidate Compound Spectra Matrix. Once the candidate atomic composition had been identified, the theoretical isotopic distribution of each compound was calculated using the statistical approach developed by Yergey.[28] The relative abundances of all isotopic combinations were calculated as follows:

$$\mathrm{RA} = \prod_{i=\text{number of elements}} A_i \qquad (1)$$

where RA indicates the abundance of an isotopic combination for the candidate molecule, and A_i indicates the abundance of elements of *i* in the molecule

$$A_i = \frac{n!}{(a!)(b!)(c!)...}(r_a)^a(r_b)^b(r_c)^c... \qquad (2)$$

$$a + b + c + ... = n \qquad (3)$$

where *n* is the number of a given element *i* in the molecule, and *a*, *b*, and *c* indicate the possible combinations of the isotopes of the element of *i*. r_a, r_b, and r_c are the isotopic compositions of *a*, *b*, and *c*, respectively.

To calculate the relative intensity of abundant isotopic combinations, the following equation was used:

$$I_i = \frac{\mathrm{RA}_i}{\max\{\mathrm{RA}\}} \times 100\% \qquad (4)$$

where I_i is the theoretical intensity of a given isotopic combination.

DOI: 10.1021/acs.est.6b01349
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

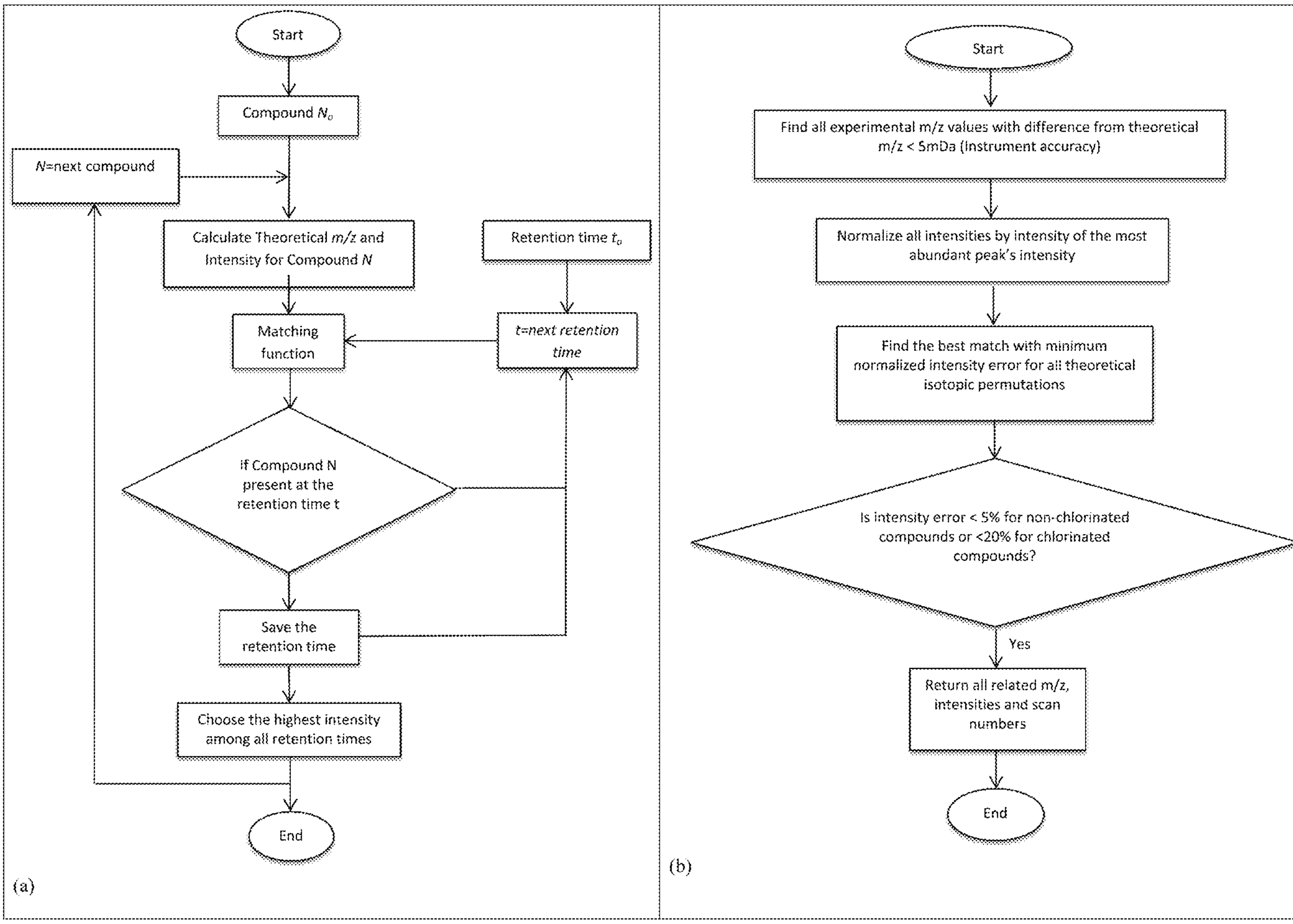


Figure 1. Flowcharts of the (a) screening algorithm and (b) matching function.

The atomic mass and isotopic compositions of the elements used in this work are shown in Table S.2. On the basis of these equations, theoretical m/z and corresponding relative intensities for the isotopic distribution of each molecule were calculated (for example, see Table S.5.). Theoretical isotopic intensities less than 4% were not considered significant for candidate matching and were excluded from the candidate library.

Masses of specific isotopic combinations can be very close and form a cluster of peaks in a mass window less than the mass accuracy of the instrument. For example, the mass difference between $^{12}C_8F_{16}{}^{35}Cl^{34}S^{16}O_3^-$ (516.8969 Da) and $^{12}C_8F_{16}{}^{37}Cl^{32}S^{16}O_3^-$ (516.8972 Da) is 0.3 mDa (more details are provided in Supporting Information, Tables S.3 and S.4, and Figures S.1 and S.2). To overcome this issue, each peak cluster was merged into a single peak using the rectangular mass window method that is independent of individual intensities.[29] An arbitrary initial m/z point was chosen and an interval was applied in a repeating manner across a desired m/z range (Figure S.1). Signals within the interval were then merged into a unique signal using eqs 5 and (6). In this work, the monoisotopic mass was chosen as the arbitrary initial m/z point and the interval was set equal to the instrument accuracy.

The total abundance of the individual peaks in each cluster was considered the abundance of the merged peak, with the molecular weight calculated from the weighted average of all peaks in the cluster.

$$\mathrm{RA}_{\mathrm{cluster}} = \sum_{i=1}^{\substack{\text{number of peaks}\\ \text{in the cluster}}} \mathrm{RA}_i \tag{5}$$

$$\mathrm{MW}_{\mathrm{cluster}} = \frac{\sum_{i=1}^{\substack{\text{number of peaks}\\ \text{in the cluster}}} \mathrm{RA}_i \times \mathrm{MW}_i}{\mathrm{RA}_{\mathrm{cluster}}} \tag{6}$$

Candidate List Screening. MassWolf[30] was found to be more compatible with MS^e mode data configuration and was chosen over the more commonly used ProteoWizard[31] (version 3.0.7414) to convert raw instrument files (MassLynx, Waters, Milford, MA) to the MATLAB compatible mzXML formats. Data were collected using multiple energy channels by the instrument, therefore, a MATLAB script beyond the preprogramed mzXML convertor function in the MATLAB bioinformatics toolbox was needed to import and search each data file.

Two scripts labeled Screening Function and Matching Function (Figure 1), respectively, were developed to find theoretical m/z values within the instrument's mass accuracy range (<5 mDa or 10 ppm for PFOS) for all candidate compounds (MATLAB functions are available in Supporting Information). In many cases, multiple experimental m/z value matches were detected for a given isotopic combination (compound) at different retention times. To identify the group or cluster with the best experimental match, the

DOI: 10.1021/acs.est.6b01349
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

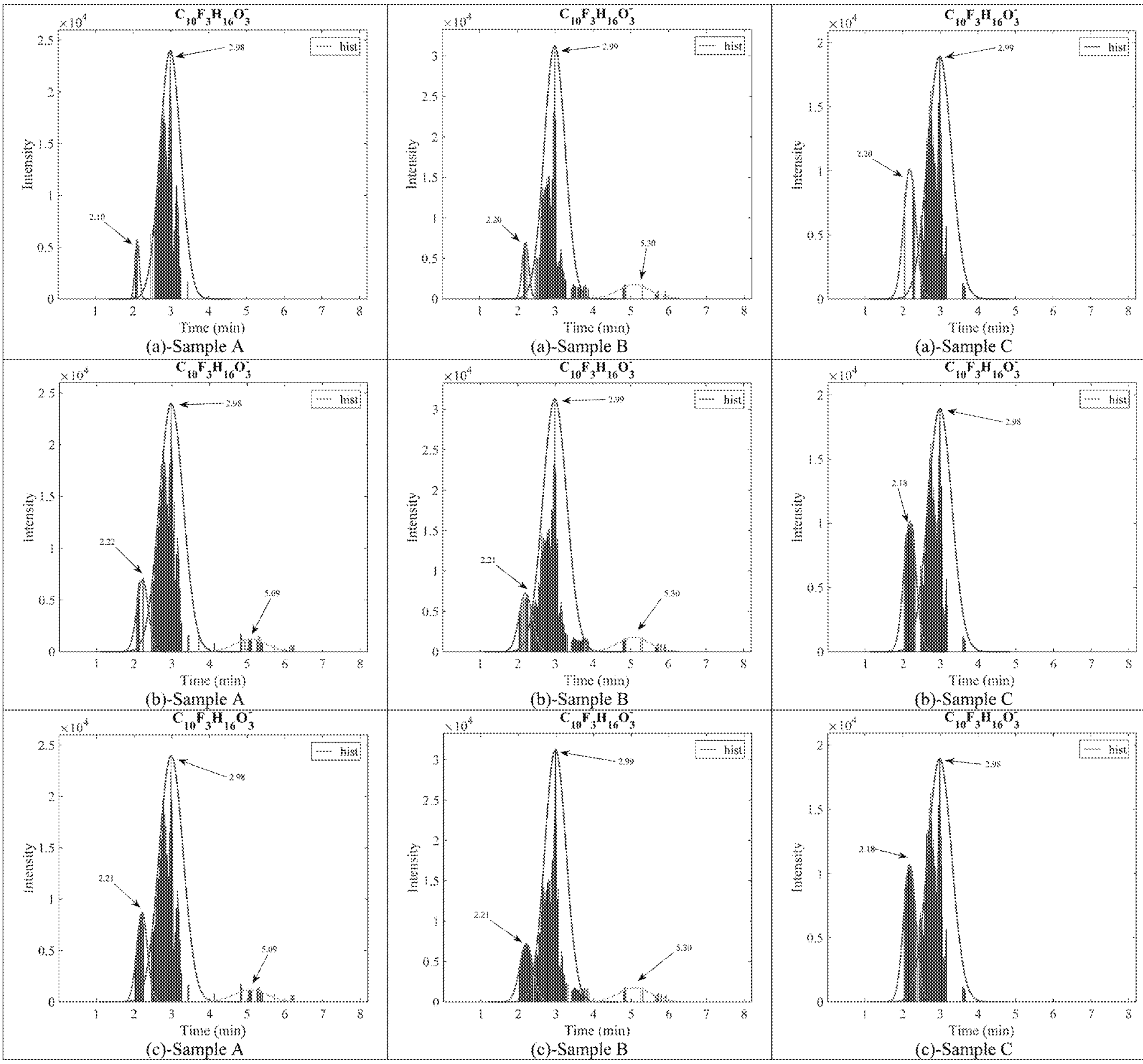


Figure 2. Intensity versus retention time for each scan time of $C_{10}F_3H_{16}O_3^-$ illustrating the peak fitting selection criteria for including candidate compound in the "found" list for samples A, B, and C. (a) Isotopic profile match for $[M - H]^-$ and presence of $[M - CO_2]^-$ at both low and high energy channels at same scan; (b) isotopic profile match for $[M - H]^-$ and presence of $[M - CO_2]^-$ in only the low energy channel of the same scan; (c) isotopic profile match for $[M - H]^-$.

intensities of all identified experimental m/z were normalized by the intensity of the highest isotopic abundances for the candidate compound potentially matching the m/z cluster. Thus, the number of potential matching mass clusters is a factor of the number of the experimental m/z values that match the highest intensity isotopic combination. The objective function below was minimized to select the best match of experimental peaks in a spectrum.

$$\mathrm{OF} = \sum_{i=1}^{\substack{\text{number of isotopic} \\ \text{combinations}}} \frac{|I_{\mathrm{exp}} - I_{\mathrm{the}}|}{\sqrt{I_{\mathrm{the}}}} \tag{7}$$

This objective function was based on the minimization of the relative intensity error of each experimental mass cluster.

Retention Time Determination. Within an analysis, there may be m/z peaks matching a candidate peak at multiple retention times. A higher intensity for a given m/z match alone does not constitute a correctly identified peak. Therefore, in the current identification method, the retention time was selected based on the isotopic pattern of the $[M - H]^-$ ion, the presence of $[M - CO_2]^-$ ion for polyfluoro carboxylic acids in both the high and low energy channels, and if the most abundant isotopic combination (highest intensity) form a reasonable chromatographic peak shape (similar to that shown in Figure 2). If histograms of at least three consecutive scan numbers resembled peak shapes through visual inspection, the cluster of histograms was called a peak. As seen in Figure 2c, this technique potentially resolved three peaks that may correspond to conformational isomers of $C_{10}F_3H_{16}O_3^-$.

DOI: 10.1021/acs.est.6b01349
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

The retention time finder algorithm employed the following criteria: select the retention time that the highest intensity occurs, and maintain an ascending trend for log K_{ow} versus retention time

The retention time of a candidate compound was determined in a three-step process as illustrated in Figure 2. After detecting the isotopic profile match, files were screened for potential retention times where $[M - CO_2/SO_3]^-$ ions were present in both the low and high energy channels (Figure 2a). If these peaks were not found, files were screened for $[M - CO_2/SO_3]^-$ ions in only the low energy channel (Figure 2b). If this match also did not occur, only isotopic profile matches were used to identify the retention time (Figure 2c).

As candidate peaks were selected using isotopic profiles, a molecular conformation was generated based on probable structures assumed above. An input SMILES (simplified molecular-input line-entry system) was then generated and entered into the U.S. Environmental Protection Agency EPI Suite v4.11 (www.epa.gov/oppt/exposure/pubs/episuitedl.htm) to calculate theoretical octanol−water partitioning coefficients (K_{ow}) of the protonated form. Several assumptions were made in generating the input SMILES for unknown compounds, including a linear molecular configuration, carboxyl and not ester functionalities when two or more oxygen atoms present, trifluoromethyl groups when three fluorine atoms present, fluorine atoms populate the alkyl end of the linear molecule, and ether groups present in the middle of the molecule.

Retention time reproducibility (0.1 min) among replicate injections and relative retention time vs log K_{ow} was then used to discriminate between noise and candidate compounds. Generally, increasing log K_{ow} results in increased retention times for a reversed-phase system. In addition, an ascending trend of log K_{ow} versus retention time relationship should be observed within a compound class. The chromatographic behavior of the labeled PFAAs added prior to the previous targeted analyses was used as a retention time/log K_{ow} trend reference for the current chromatographic method. Although pK_a can affect the retention of acidic/basic species in reversed-phase LC systems, the calculated pK_as for the species observed did not vary using the above assumptions (SPARC,[33] Table S.7) and were excluded from the selection criteria. Standards and blanks processed as part of the targeted analysis were used to remove background.

Polymers with nonfluorinated carbon backbones and perfluoroalkyl side chains might be potential precursors of PFAAs.[34] Hence, standard solutions of PFAAs and PFAS were analyzed to ensure that the detected compounds were not the result of impurities or potential in-source products from the isotopically labeled PFAAs and PFAS added to the fish extracts. Spurious m/z within an instrument run is not uncommon when searching for low level contaminants in raw spectra. In this study, fish were extracted in triplicate to evaluate sample-specific contamination, and each extract was injected 3 times to ensure detected m/z were not the result of instrument noise.

RESULTS

Method Validation Using PFAA and PFAS Standards. Calibration solutions containing ^{13}C and native PFAAs were used to test the algorithm (Table S.6 and Figure S.3). The maximum absolute intensity error of the isotope model and maximum measured mass error were observed for perfluorohexanesulfonate at 9.25% and ~4.5 mDa, respectively. The average absolute error of the isotope model was 1.48%. The positive relationship between the retention time and log K_{ow} (derived from EPI Suite[32]) is shown in Figure S.4a. Expected isotopic patterns for the $[M - H]^-$, and the $[M - CO_2]^-$ ions for PFAAs and $[M - SO_3]^-$ ion for PFBS at the same retention time in the low energy channel were also identified.

The instrumental mass accuracy threshold was set at 5 mDa (10 ppm for PFOS) based on measured mass error of standard solutions. A maximum intensity error among isotopic combinations for nonchlorinated and chlorinated compounds was set to 5% and 20%, respectively. The higher acceptable intensity error for chlorinated compounds was used since they have more diverse isotopic combinations in each isotopic distribution. Features with intensity errors above these thresholds were excluded.

Application to Biological Matrices. Archived data files from three different 2011 Lake Michigan lake trout extracts that had been analyzed in triplicate were examined with concurrent laboratory blanks. If a candidate satisfied the selection criteria mentioned above, and occurred within a 0.1 min retention time window for each replicate injection, the compound was assumed to be present in the sample. The experimental retention time was determined using histograms similar to Figures 2a, b, and c based on the average retention times and intensities. As expected when analyzing a complex biological matrix, elevated noise levels were observed so an absolute minimum ion intensity of 500 for the most abundant isotopic mass was applied to improve the signal-to-noise ratio of candidate molecules. Identified candidate retention times versus log K_{ow} were plotted for each sample (Figure S.3.). If multiple matching isotopic profiles were present for a compound, the retention time that made the most sense chromatographically was selected.

Identified Molecular Formula. The postulated molecular formulas can be classified into eight groups: monofluoroalkyl carboxylic acids (MFCAs), trifluoroalkyl carboxylic acids (TrFCAs), tetrafluoroalkyl carboxylic acids (TeFCAs), pentafluorodecanoic acid (PeFDA), hexafluorodecanoic acid (HFCA), monofluoroalkyl ether carboxylic acids (MFECA), trifluoroalkyl ether carboxylic acids (TrFECAs), and polyfluoroalkyl sulfonate (PoFAS). Structures of these molecular formulas are presented in Table S.8.

The **monofluoroalkyl carboxylic acids ($C_nFH_{2n-1}O_2$)** molecular formulas consisted of six carbon lengths (C_5−C_{10}). Retention time differences for the parent ion were not significant (within 0.1 min) among the three samples for the majority of molecular formulas (Table S.7). The current analytical method typically generates a $[M - CO_2]^-$ fragment in the low and high energy channels for the linear perfluoroalkyl carboxylic acids. The $[M - CO_2]^-$ fragment for $C_7FH_{12}O_2^-$ was observed at same retention time in the low energy channel in replicate samples confirming the presence of a carboxyl groups. However, carboxyl group for other molecular formulas was not observed in all triplicate injections for $C_5FH_8O_2^-$, $C_6FH_{10}O_2^-$, and $C_8FH_{14}O_2^-$, and it was only present in triplicate injections for $C_9FH_{16}O_2^-$ and $C_{10}FH_{18}O_2^-$.

5-Fluoropentanoic acid from this class of compounds was commercially available, and purchased from Enamine Ltd. (Kyiv, Ukraine). The standard spectrum of this compound is shown in Figure S.6. $[M - H]^-$ and $[M - F - H]^-$ were the major fragments and compared well with the candidate identified in lake trout shown in Table S.6. The profile relative intensities (normalized by $[M - H]^-$ and retention time of the

DOI: 10.1021/acs.est.6b01349
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

neat standard differed by less than 2 standard deviations, respectively, from the mean of the candidate identified in the three trout extractions.

Monofluorinated fatty acids have been observed in South Africa plants but are rare in nature and significantly larger than the chain lengths observed here suggesting an alternative source to the lake trout.[35]

Trifluoroalkyl carboxylic acids ($C_nF_3H_{2n-3}O_2$) were identified for seven trifluorinated carbon chain lengths (C_4–C_{10}). The $[M - CO_2]^-$ ion was present at same retention time in the low and high energy channels for the majority of molecular formulas. The proposed $C_4F_3H_4O_2^-$ and $C_5F_3H_6O_2^-$ isomers do not display an increasing retention time vs log K_{ow} trend (Figure S.3) suggesting assigned molecular formulas of these masses should be viewed with caution. The remaining (C_6–C_{10}) TrFCAs follow the expected ascending trend of retention time vs log K_{ow} consistent with an increasing carbon homologue distribution. The only commercially available compound from this class was 4,4,4-trifluorobutanoic acid, which was purchased from Sigma-Aldrich (St. Louis, MO). The spectrum of this species is shown in Figure S.7. $[M - H]^-$ and $[M - F - H]^-$ are the primary fragments, with lesser contributions from $[M - F_2 - H_2]^-$ and $[M - F_3 - H_3]^-$. Unlike the species isolated in the lake trout, $[M - CO_2]^-$ was a very minor component of the standard spectrum. The differences in spectral profiles suggest the candidate and standard may have different (possibly isobutyl) molecular conformations even though the retention time observed for the standard was within two standard deviations of the triplicate extraction mean. The current method was not optimized for these classes of compounds and coelution of conformational isomers is not unexpected.

Two species (C_9 and C_{10}) of **tetrafluoroalkyl carboxylic acids ($C_nF_4H_{2n-4}O_2$)** were detected. The $[M - CO_2]^-$ ion was not observed and this class of molecular formula was not present in all of the replicate extractions. For example, $C_9F_4H_{13}O_2^-$ was not present in all of the triplicate injections of sample C, and $C_{10}F_4H_{15}O_2^-$ was present in all triplicate injections of sample B. These features were detected at low levels and the absence of the C_9 and C_{10} in a replicate extraction may be due to detection limitations for this chemical formula. An alternative configuration may be the $-C(CO_2)F-$ conformation or fluorine saturated carbons adjacent to the carboxyl group, analogous to chlorinated and nonchlorinated polyfluorinated ether sulfonates.[27] Liu et al.[16] observed a $C_9F_{12}H_7SO_3^-$ peak in Chinese wastewater and postulated a structure resulting from repeating $-CF_2$ CH_2-, or $-CHF-$ subunits employed in metal siding and chemical processing equipment. It is possible that this class of molecular formula is formed in a similar way.[36]

One **pentafluorodecanoic acid ($C_{10}F_5H_{15}O_2$)** was detected. Although the molecular ion was present in all samples, the $[M - CO_2]^-$ ion was only observed in the low energy channel of sample C. Again, sensitivity issues may limit the detection of this chemical formula or this mass spectral feature may be an anomaly in this extraction.

One **hexafluorodecanoic acid ($C_{10}F_6H_{14}O_2$)** was detected. The $C_{10}F_6H_{13}O_2^-$ candidate was only present in sample B. Similar to the tetrafluoroalkyl carboxylic acids, the conformation of this class of molecular formula may consist of fluorine saturated carbons adjacent to an ether or carboxyl group. Alternatively, the industrial blend Zonyl FSN (Dupont) was recently found to contain a series of polyfluoroalkyl ethoxylates, $F(CF_2)_x(CH_2CH_2O)_y$ with x and y values up to 18 and 50, respectively, and half-lives greater than 48 days for the shorter chain analogues.[37] The C_9 and C_{10} homologous for the tetrafluoroalkyl carboxylic acids and hexafluorodecanoic acid could be the result of oxidation of the ethoxylate through environmental or metabolic processes. However, the conformation presented by Frömel and Knepper[37] does not include partially fluorinated alkyl chains with greater than two methylene groups suggesting a repeating $-(CF_2)_x(CH_2)_x-$. Regardless, the similar carbon chain length (C_9 and C_{10}) for the F4 and F6 acids indicates similar feedstocks and sources. Alkylphenol ethoxylates have been widely used as surface tension modifiers, with the nonylphenol being the dominant species observed in environmental samples.[38] The C_9 and C_{10} tetra- and hexafluoroethoxylate would likely have similar, if not enhanced, surface modifying properties and bioaccumulation potential as the nonylphenol ethoxylates.[39,40] Once oxidized, this form would have a hybrid, free fatty acid/perfluoroalkyl acid activity in biota.

The **monofluoroalkyl ($C_nFH_{2n-1}O_3$)** and **trifluoroalkyl ether carboxylic acids ($C_nF_3H_{2n-3}O_3$)** contained six (C_5–C_{10}) and five molecular formulas, respectively. The $[M - CO_2]^-$ fragment was observed for most of the identified peaks in these classes within the 0.1 min retention time window for both the low and high energy channels. $C_7FH_{12}O_3^-$ was not present in all injections of samples A and B, but the $C_8FH_{14}O_3^-$ and $C_{10}FH_{18}O_3^-$ homologues exhibited the least retention time variance among samples.

The **polyfluoroalkyl sulfonates ($C_cF_fH_{2c-f+2}SO_3$)** include $C_6F_2H_{11}SO_3^-$ and $C_8F_8H_9SO_3^-$. An apparent $[M - SO_3]-$ fragment of $C_6F_2H_{11}SO_3^-$ was detectable in the low energy channel in samples A and B. The loss of SO_3^- was not observed for the linear perfluorinated sulfonates suggesting a nonlinear conformation or the presence of an ether subunit. Ruan et al.[27] observed the loss of a $C_2F_4SO_3^-$ fragment of 8:2 Cl-PFAES (chlorinated polyfluorinated ether sulfonate) using a triple quadrupole. The even number of fluorine atoms in these molecules may indicate an alternating $-(CF_2)_x(CH_2)_y-$ alkyl configuration in conjunction with ether sulfonate moieties.

Although the linear configuration was assumed when calculating candidate log K_{ow}s, a variety of isomers may be present, identified, or coelute in the LC gradient. In Figure 2, there appears to be 2–3 resolved isomers for $C_{10}F_3H_{16}O_3^-$ identified in the triplicate extraction. In samples A, B, and C, the two highest intensity peaks elute around 2.10–2.2 min and 2.98–2.99 min, respectively. In the LC gradient used, branched isomers typically elute before the linear form and it is well documented that the linear perfluoroalkyl isomers dominate the isomeric signature in biota.[1] Extrapolating this to the isomeric pattern observed for $C_{10}F_3H_{16}O_3^-$ suggests the detection of branched and linear isomers having this molecular formula in lake trout.

Except for the low molecular weight trifluoroalkyl carboxylic acids in samples A and C, most of the candidate chemical formulas showed a positive log K_{ow}/retention time relationship expected in a reverse phase liquid chromatographic system. The average relative standard deviation (RSD) of the intensities from triplicate injections of the triplicate extractions was 19% for all species detected (ranging 3.0% and 47% for $C_5F_3H_6O_2^-$ and $C_6F_3H_8O_3^-$, respectively). This value is within expected errors associated with replicate extractions, suggesting the observed intensities were not spurious signals within a run.

DOI: 10.1021/acs.est.6b01349
Environ. Sci. Technol. XXXX, XXX, XXX–XXX

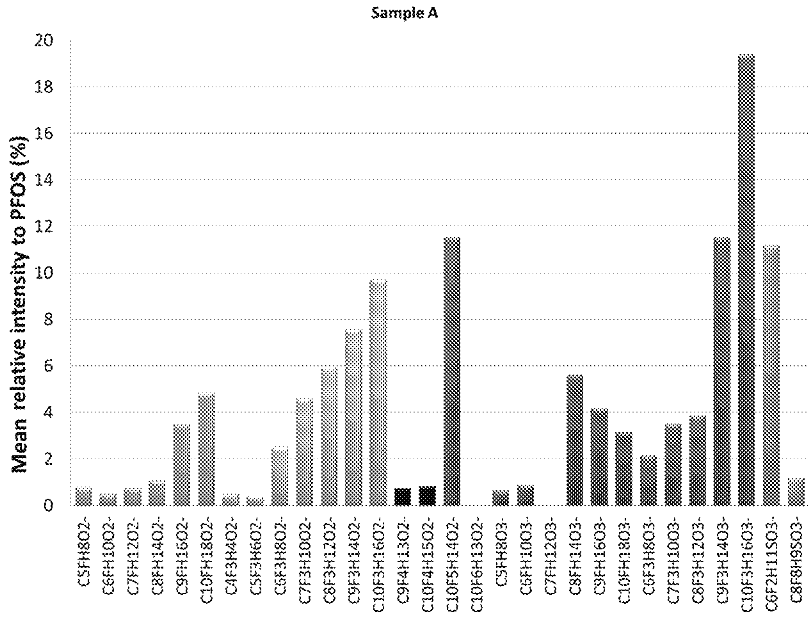


Figure 3. Comparison of mean relative intensities of identified compounds to PFOS for Sample A. Color bars: orange, monofluoroalkyl carboxylic acids; green, trifluoroalkyl carboxylic acids; black, tetrafluoroalkyl carboxylic acids; blue, pentafluorodecanoic acid; yellow, hexafluorodecanoic acid; red, monofluoroalkyl ether carboxylic acids; purple, trifluoroalkyl ether carboxylic acids; turquoise, polyfluoroalkyl sulfonate.

In general, relative intensities within a compound class increased with carbon number with the maximum intensities typically found for the C_9 and C_{10} carbon lengths (Figures 3 and S.4). The intensities of these peaks were the same order of magnitude as PFOS, suggesting these compounds are present at significant levels in lake trout. Increasing intensity with chain length and K_{ow} (Table S.7, Figure S.4) also suggest physicochemical properties govern the bioaccumulation of these chemicals in the lake trout.

The majority of the molecular formulas identified may be associated with recent manufacturing activities and emitted to the environment in industrial streams or may be byproducts of the transition from long chain perfluorinated chemicals to polyfluorinated ethers, ethoxylates, and sulfonate ethers. Alternatively, these chemical formulas may be byproducts or impurities of other fluorinated products. Synthesizing neat standards to confirm the presence of these chemical formulas would be the next step but is only possible because the algorithm has effectively identified several species out of the thousands of possible compounds.

■ METHOD IMPLICATIONS

Monitoring programs like GLFMSP are increasingly tasked with identifying the next chemicals of concern. This new direction must not, however, sacrifice current targeted chemical monitoring. Fortunately, newly available scanning high resolution mass spectrometers, such as the Waters Xevo G2 QToF employed in this study, allow for full scan, accurate mass measurements, with the linear ranges and sensitivities sufficient to quantify "monitored" compounds (PFAAs and PFASs) at environmentally relevant levels.[15] The method qualified here illustrates the added value of archiving the full scan data generated by targeted analyses methods. The screening workflow presented was developed using a readily available format managing program, open source file conversion software, and MATLAB scripts, allowing for the universal application of this type of analyses to full scan accurate mass data. Initiating the method requires generating a candidate compound list and converting high-resolution mass spectral data. With some forethought on the compound classes of interest, boundary conditions can be set based on the analytical methodology used, and may allow for the identification of alternative conformations such as rings, units of unsaturation, and mixed halogenated signatures (including Cl, Br). Source-type studies[13,16,21] can be used to frame the boundary conditions (candidate structural framework), without limiting candidate subsets with nontraditional structural conformations. As long as the isotopic profiles of the compounds of interest can be calculated, this algorithm can be used. In some cases, molecular ions are not detectable, and the current searching algorithm could be adapted with additional modules incorporating known neutral loss fragmentation patterns.

The use of archived data files mitigates costly tissue re-extractions and instrument time through the generation of a database of potential unknowns, including undiscovered chemicals of concern that can be mined using this type of algorithm. Injection and recovery standards applied in the targeted analysis provide retention time, mass accuracy, and reproducibility references for each data file. Known sample-to-sample variability in measurement accuracy, reference parameters to assess physicochemical properties (log K_{ow}) of identified species, and reproducibility among replicate injections and extractions provides a comprehensive qualification workflow to sort through real and spurious spectral features.

In this study, 30 out of a possible 3570 polyfluoroalkyl candidate compounds were observed in lake trout from Lake Michigan, and, of these, only two were available commercially. Each compound matched their theoretical isotope model with a small error (5% absolute intensity error). In addition to isotopic profile, most of the carboxylated candidates had the $[M - CO_2]^-$ ion present in the low energy channel of the same scan and the closest scan of the high energy channel. In specific cases, the absence of the $[M - CO_2]^-$ fragment may suggest a different conformation. The above parameters were observed in three replicate extractions (triplicate injection) of a lake trout collected from Lake Michigan. This workflow can be easily expanded to other candidate lists, ionization modes, and fragmentation sequences,[21] affording a cost-effective means to catalog the chemical fingerprints of Great Lakes trout without additional analyses that deplete finite sample tissue stores and add the expense of additional extraction.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.6b01349.

> Brief description of the sample preparation and instrumental methods, table outlining the composition of the candidate matrix, illustration of the rectangular window method, calculated pK_as for identified molecular formulas, calibration results table and figure for PFAA analysis, retention time vs log K_{ow} plots for per- and polyfluorinated compounds identified, plots illustrating the amount of each identified molecular formula relative to PFOS, and a file containing the MATLAB code and an example of results used in this manuscript (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*Tel: 202-368-6926; e-mail: bcrimmin@clarkson.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The U.S. Environmental Protection Agency Great Lakes Fish Monitoring and Surveillance Program supported this project under contract GL-96594201-1. We also wish to thank the Program Manager Elizabeth Murphy and many people who assisted in sample collection and processing. We are also thankful of Dr. Mahesh Banavar of Clarkson University for assistance in applying the rectangular mass window method. Although the research described in this article has been funded wholly or in part by the United States Environmental Protection Agency, it has not been subjected to the Agency's required peer and policy review and therefore, does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

■ REFERENCES

(1) Houde, M.; De Silva, A. O.; Muir, D. C. G.; Letcher, R. J. Monitoring of Perfluorinated Compounds in Aquatic Biota: An Updated Review. *Environ. Sci. Technol.* **2011**, *45* (19), 7962–7973.

(2) Lindstrom, A. B.; Strynar, M. J.; Libelo, E. L. Polyfluorinated Compounds: Past, Present, and Future. *Environ. Sci. Technol.* **2011**, *45* (19), 7954–7961.

(3) UNEP. *The 9 New POPs*. Stockholm Convention on Persistent Organic Pollutants (POPs); 2009.

(4) U.S. EPA. *Emerging Contaminants – Perfluorooctane Sulfonate (PFOS) and Perfluorooctanoic Acid (PFOA)*; Washington, DC, 2014.

(5) ECHA. *European Council for High Ability*; 2014.

(6) Wang, Z.; Cousins, I. T.; Scheringer, M.; Buck, R. C.; Hungerbühler, K. Global emission inventories for C4–C14 perfluoroalkyl carboxylic acid (PFCA) homologues from 1951 to 2030, Part I: production and emissions from quantifiable sources. *Environ. Int.* **2014**, *70*, 62–75.

(7) Heydebreck, F.; Tang, J.; Xie, Z.; Ebinghaus, R. Alternative and Legacy Perfluoroalkyl Substances: Differences between European and Chinese River/Estuary Systems. *Environ. Sci. Technol.* **2015**, *49* (14), 8386–8395.

(8) Martin, J. W.; Kannan, K.; Berger, U.; Voogt, P. D.; Field, J.; Franklin, J.; Giesy, J. P.; Harner, T.; Muir, D. C. G.; Scott, B.; Kaiser, M.; Järnberg, U.; Jones, K. C.; Mabury, S. A.; Schroeder, H.; Simcik, M.; Sottani, C.; Bavel, B. V.; Kärrman, A.; Lindström, G.; Leeuwen, S. V. Peer Reviewed: Analytical Challenges Hamper Perfluoroalkyl Research. *Environ. Sci. Technol.* **2004**, *38* (13), 248A–255A.

(9) Strynar, M. J.; Lindstrom, A. B. Perfluorinated Compounds in House Dust from Ohio and North Carolina, USA. *Environ. Sci. Technol.* **2008**, *42* (10), 3751–3756.

(10) Kannan, K.; Corsolini, S.; Falandysz, J.; Fillmann, G.; Kumar, K. S.; Loganathan, B. G.; Mohd, M. A.; Olivero, J.; Wouwe, N. V.; Yang, J. H.; Aldous, K. M. Perfluorooctanesulfonate and Related Fluorochemicals in Human Blood from Several Countries. *Environ. Sci. Technol.* **2004**, *38* (17), 4489–4495.

(11) Posner, S.; Roos, S.; Brunn Poulsen, P.; Ólína Jörundsdottir, H. *Per- and Polyfluorinated Substances in the Nordic Countries: Use, Occurrence and Toxicology*, 1st ed.; Nordic Council of Ministers: Copenhagen, 2013.

(12) Yamamoto, A.; Hisatomi, H.; Ando, T.; Takemine, S.; Terao, T.; Tojo, T.; Yagi, M.; Ono, D.; Kawasaki, H.; Arakawa, R. Use of high-resolution mass spectrometry to identify precursors and biodegradation products of perfluorinated and polyfluorinated compounds in end-user products. *Anal. Bioanal. Chem.* **2014**, *406* (19), 4745–4755.

(13) Myers, A. L.; Jobst, K. J.; Mabury, S. A.; Reiner, E. J. Using mass defect plots as a discovery tool to identify novel fluoropolymer thermal decomposition products. *J. Mass Spectrom.* **2014**, *49* (4), 291–296.

(14) van Leeuwen, S. P. J.; Kärrman, A.; van Bavel, B.; de Boer, J.; Lindström, G. Struggle for Quality in Determination of Perfluorinated Contaminants in Environmental and Human Samples. *Environ. Sci. Technol.* **2006**, *40* (24), 7854–7860.

(15) Crimmins, B.; Xia, X.; Hopke, P.; Holsen, T. A targeted/non-targeted screening method for perfluoroalkyl carboxylic acids and sulfonates in whole fish using quadrupole time-of-flight mass spectrometry and MSe. *Anal. Bioanal. Chem.* **2014**, *406* (5), 1471–1480.

(16) Liu, Y.; Pereira, A. D. S.; Martin, J. W. Discovery of C5–C17 Poly- and Perfluoroalkyl Substances in Water by In-Line SPE-HPLC-Orbitrap with In-Source Fragmentation Flagging. *Anal. Chem.* **2015**, *87* (8), 4260–4268.

(17) Sleno, L. The use of mass defect in modern mass spectrometry. *J. Mass Spectrom.* **2012**, *47* (2), 226–236.

(18) Kendrick, E. A Mass Scale Based on CH2 = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Anal. Chem.* **1963**, *35* (13), 2146–2154.

(19) Barzen-Hanson, K. A.; Field, J. A. Discovery and Implications of C2 and C3 Perfluoroalkyl Sulfonates in Aqueous Film-Forming Foams and Groundwater. *Environ. Sci. Technol. Lett.* **2015**, *2* (4), 95–99.

(20) Hilton, D. C. Automated screening for hazardous components in complex mixtures based on functional characteristics identifiable in GC×GC-TOF-MS data. *Curr. Trends Mass Spectrom.* **2007**, *0*, 28–34.

(21) Strynar, M.; Dagnino, S.; McMahen, R.; Liang, S.; Lindstrom, A.; Andersen, E.; McMillan, L.; Thurman, M.; Ferrer, I.; Ball, C. Identification of Novel Perfluoroalkyl Ether Carboxylic Acids (PFECAs) and Sulfonic Acids (PFESAs) in Natural Waters Using Accurate Mass Time-of-Flight Mass Spectrometry (TOFMS). *Environ. Sci. Technol.* **2015**, *49* (19), 11622–11630.

(22) Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I. M.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N. S.; Bletsou, A.; Zwiener, C.; Ibáñez, M.; Portolés, T.; de Boer, R.; Reid, M. J.; Onghena, M.; Kunkel, U.; Schulz, W.; Guillon, A.; Noyon, N.; Leroy, G.; Bados, P.; Bogialli, S.; Stipaničev, D.; Rostkowski, P.; Hollender, J. Non-target screening with high resolution mass spectrometry: Critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem.* **2015**, *407* (21), 6237–6255.

(23) Chang, F.; Pagano, J. J.; Crimmins, B. S.; Milligan, M. S.; Xia, X.; Hopke, P. K.; Holsen, T. M. Temporal trends of polychlorinated biphenyls and organochlorine pesticides in Great Lakes fish, 1999–2009. *Sci. Total Environ.* **2012**, *439*, 284–290.

(24) Omara, M.; Crimmins, B. S.; Back, R. C.; Hopke, P. K.; Chang, F.-C.; Holsen, T. M. Mercury biomagnification and contemporary food web dynamics in lakes Superior and Huron. *J. Great Lakes Res.* **2015**, *41* (2), 473–483.

(25) Xia, X.; Hopke, P. K.; Crimmins, B. S.; Pagano, J. J.; Milligan, M. S.; Holsen, T. M. Toxaphene trends in the Great Lakes fish. *J. Great Lakes Res.* **2012**, *38* (1), 31–38.

(26) Crimmins, B. S.; Pagano, J. J.; Milligan, M. S.; Holsen, T. M. Environmental Mass Spectrometry in the North American Great Lakes Fish Monitoring and Surveillance Program. *Aust. J. Chem.* **2013**, *66* (7), 798–806.

(27) Ruan, T.; Lin, Y.; Wang, T.; Liu, R.; Jiang, G. Identification of Novel Polyfluorinated Ether Sulfonates as PFOS Alternatives in Municipal Sewage Sludge in China. *Environ. Sci. Technol.* **2015**, *49* (11), 6519–6527.

(28) Yergey, J. A. A general approach to calculating isotopic distributions for mass spectrometry. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *52* (2–3), 337–349.

(29) Spanias, A. *Digital Signal Processing: An Interactive Approach*, 2nd ed.; Lulu Publishers: Raleigh, NC, 2014.

(30) Tasman, N.; Philosof, R. S.; Tchekhovskoi, D. *MassWolf*, 4.3.1; 2009.

(31) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–920.

(32) U.S. EPA. *Estimation Programs Interface Suite for Microsoft® Windows, v 4.11*; Washington, DC, USA, 2015.

(33) Hilal, S. H.; Karickhoff, S. W.; Carreira, L. A. A Rigorous Test for SPARC's Chemical Reactivity Models: Estimation of More Than 4300 Ionization pKa's. *Quant. Struct.-Act. Relat.* **1995**, *14* (4), 348–355.

(34) OECD. *OECD/UNEP Global PFC Group Synthesis Paper on Per- and Polyfluorinated Chemicals (PFCs), Environment, Health and Safety, Environment Directorate, OECD*; Organisation for Economic Co-operation and Development: Paris, 2013.

(35) Dembitsky, V. M.; Srebnik, M. Natural halogenated fatty acids: their analogues and derivatives. *Prog. Lipid Res.* **2002**, *41* (4), 315–367.

(36) Kricheldorf, H. R.; Nuyken, O.; Swift, G. In *Handbook of Polymer Synthesis*, 2nd ed; Marcel Dekker: New York, 2005.

(37) Frömel, T.; Knepper, T. P. Fluorotelomer ethoxylates: Sources of highly fluorinated environmental contaminants part I: Biotransformation. *Chemosphere* **2010**, *80*, 1387–1392.

(38) Renner, R. European Ban on Surfactants Triggers Transatlantic Debate. *Environ. Sci. Technol.* **1997**, *31*, 316A–320A.

(39) Rice, C. P.; Schmitz-Afonso, I.; Loyo-Rosales, J. E.; Link, E.; Thoma, R.; Fay, L.; Altfater, D.; Camp, M. J. Alkylphenol and alkylphenol-ethoxylates in carp, water and sediments from Cuyahoga River, Ohio. *Environ. Sci. Technol.* **2003**, *37*, 3747–3754.

(40) Furdui, V. I.; Stock, N. L.; Ellis, D. A.; Butt, C. M.; Whittle, D. M.; Crozier, P. W.; Reiner, E. J.; Muir, D. C. G.; Mabury, S. A. Spatial Distribution of Perfluoroalkyl Contaminants in Lake Trout from the Great Lakes. *Environ. Sci. Technol.* **2007**, *41* (5), 1554–1559.

DOI: 10.1021/acs.est.6b01349
Environ. Sci. Technol. XXXX, XXX, XXX–XXX